LA-UR-89-1989

CONF-8906159--2
Received by OSTI

JUL 07 1989

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

LA-UR--89-1989

DE89 014252

TITLE The Human Telomere

AUTHOR(S) Robert K. Moyzis

SUBMITTED TO Proceedings of the 6th Conversation in the Discipline Biomolecular Stereodynamics

DISCLAIMER

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

The Human Telomere

Robert K. Moyzis

Los Alamos National Laboratory, Los Alamos, NM 87545

Abstract

An ultimate goal of human genetics is the generation of a complete physical and "functional" map of the human genome. Twenty-five percent of human DNA, however, consists of repetitive DNA sequences. These repetitive DNA sequences are thought to arise by many mechanisms, from direct sequence amplification by the unequal recombination of homologous DNA regions to the reverse flow of genetic information. A general outline of the chromosomal organization of these repetitive sequences will be discussed. Our working hypothesis is that certain classes of human repetitive DNA sequences "encode" the information necessary for defining long-range genomic structure. Evidence will be presented that the first goal of this research, the identification and cloning of the human telomere, has been achieved.

A human repetitive DNA library was constructed from randomly sheared, reassociated, and oligo(G·C)-tailed DNA, a method that minimizes the potential loss of sequences devoid of a given restriction enzyme site. Sequences too large to clone efficiently in cosmid or λ vectors, such as centromeric repeats, or telomeric sequences with an end incompatible for cloning, should be present in this library. In order to isolate highly conserved repetitive DNA sequences, this library was screened with radiolabeled hamster Cot50 repetitive DNA. Two clones, containing tandem arrays of the sequence (TTAGGG), were isolated by this method. The identity of this sequence to that reported

previously for the DNA present at trypanosome telomeres suggested that the $(TTAGGG)_n$ sequence may be present at human telomeres as well.

In situ hybridization to human metaphase chromosomes localized the major clusters of this sequence at the telomeres of all chromosomes. Quantitative hybridizations to flow-sorted human chromosomes indicated that similar amounts of this sequence are present on each chromosome, regardless of chromosome length. Bal-31 nuclease digestion experiments indicate that the major clusters of this sequence (approximately 250 to 1000 hexamers) are near or directly at the chromosome termini. Using synthetic oligodeoxynucleotides, conditions were determined under which the human telomere sequence $(TTAGGG)_n$ does not cross-hybridize to other known telomere sequences. Using these conditions for in situ hybridization, biotinylated human telomere oligomers were hybridized to metaphase spreads from 91 different species, including representative orders of bony fish, reptiles, amphibians, birds, and mammals. In all 91 species, hybridization to the telomeres of all chromosomes was observed, regardless of chromosome number or size. The conservation of the $(TTAGGG)_n$ sequence and its telomeric location, in species thought to share a common ancestor over 400 million years ago, strongly suggests this sequence is the functional vertebrate telomere.

The human genome contains a variety of DNA sequences present in multiple copies (1). These repetitive DNA sequences are thought to arise by many mechanisms, from direct sequence amplification by the unequal recombination of homologous DNA regions to the reverse flow of genetic information (2). While it is likely that some of these repetitive DNA sequences influence the structure and function of the human genome, little experimental evidence supports this idea at present. We reasoned, however, that evolutionary conservation of a particular repetitive DNA sequence family might be expected if the sequencce is essential to cellular function. Sequences representing centromeres, telomeres and "functional" chromosome domains would be likely candidates for such conserved repetitive DNA.

A telomere is functionally defined as a region of DNA at the molecular end of a linear chromosome that is required for replication and stability of the chromosome (3). All known eukaryotic telomeres consist of simple repeated sequences of G- and C-rich complementary strands, with the general structure $(T/A)_m(G)_n$ (3,4). The G-rich DNA strand, oriented 5'→3' towards the chromosome end, is synthesized by an RNA-dependent telomerase activity in Tetrahymena (5,6,7,8) and Oxytricha (9). Frequent recombination occurs during telomere formation in yeast genomic and Tetrahymena mitochondrial DNA, predicted by models of recombination-mediated telomere replication (10,11). Either telomerase or recombination models for telomere replication explain the stability of the basic repeating sequence, yet infrequent evolutionary change in the telomere sequence could occur with either replication method. In addition to their role in chromosome replication, functional telomeric DNA sequences are believed to confer stability to chromosomes, preventing the end-to-end fusions and DNA degradation normally observed following breakage of chromosomes by X-irradiation or physical rupture (3). Our evidence that the

functional human telomere has been identified and cloned is summarized below.

A search for highly conserved repetitive DNA sequences was initiated, utilizing a human recombinant repetitive DNA library (pHuR library, plasmid Human Repeat) (12). This library was constructed from randomly sheared and reassociated DNA, a method that minimizes the potential loss of sequences, such as centromeric repetitive DNA arrays, that are devoid of a given restriction enzyme site (13). Likewise, the ends of linear DNA molecules will, by definition, be unclonable following restriction enzyme digestion, but should be represented in this library if they consist of repetitive DNA arrays. Hybridization with a ^{32}P-labeled human $C_o t$ 50 repetitive DNA probe indicated that at least 95% of the plasmid clones in this library contain human repetitive DNA inserts. A portion of this library was screened using ^{32}P-labeled hamster $C_o t$ 50 repetitive DNA as a probe. Under standard hybridization conditions (0.95 M Na^+, 68°C), 20% of the bacterial colonies gave positive signals. Many of these clones contained Alu repetitive sequence elements, known to be found in high abundance in mammalian genomes (2). When the stringency of the hybridization conditions was increased to 0.95M Na^+, 80°C, 0.5% of the colonies still gave a strong or moderate hybridization signal. No detectable hybridization signal to any bacterial colony was observed when hybridization was conducted in 0.17 M Na^+ at 80°C. Six plasmid clones that still produced strong high stringency hybridization signals with hamster repetitive DNA were isolated for further analysis. Four of these recombinants contained small DNA inserts (39 to 48 nucleotides), each containing a variation of the same alternating (dG–dT)·(dA–dC) sequence. This sequence, with the capacity to form the alternative Z-DNA configuration, is known to be ubiquitously interspersed in eukarytoic genomes, and to be highly conserved (14). The other two clones, designated pHuR 93 and pHuR 143,

consisted of 40 and 43 copies, respectively, of highly conserved tandem arrays of the hexanucleotide sequence TTAGGG (12). This hexanucleotide sequence is identical to the hexanucleotide sequence known to be at the telomeres of trypanosome chromosomes (15,16).

Clone pHuR 93 DNA was used to determine the chromosomal distribution of the $(TTAGGG)_n$ sequence in the human genome. Quantitative slot blot analysis, using flow sorted human chromosomes (13), indicates that similar amounts of this repetitive DNA sequence are present on each human chromosome, regardless of the absolute chromosome length (12). This pattern of hybridization contrasts with that observed for other families of human tandem repetitive DNA sequences, which are localized to distinct chromosomes (13) or interspersed repeat families, such as Alu sequences (2), that give signals proportional to chromosome length. Estimates of the amount of $(TTAGGG)_n$ sequences present in the human genome, determined from both quantitative hybridization analysis and the frequency of this sequence in the pHuR library indicate that 3000 to 12,000 base pairs (500 to 2000 hexamers) are present on each human chromosome.

The genomic location of the $(TTAGGG)_n$ sequences was further characterized by fluorescent in situ hybridization (Fig. 1). In order to better control the in situ hybridization conditions, seven-mers of the hexamers (CCGTTA) and (TAACCC) were synthesized and end labeled with biotin-11-dCTP. A mixture of these two probes was hybridized to denatured human metaphase chromosomes in situ, and fluorescein-labeled avidin was used to detect the biotinylated DNA (12,13). While either strand alone gives observable hybridization signals, the fluorescent intensity increases by mixing the two, presumably due to out of register concatenation. Fluorescent signals were observed at the telomeres of all human chromosomes. As shown in Fig. 1, about 80-90% of the telomeres are clearly labeled in most metaphases. Some telomeres have very faint

hybridization, but the intensity of label appears to be a random variation. The fluorescent label is at the very end of prometaphase chromosomes but as the chromosomes condense, counterstained chromosomal DNA can be seen beyond the labeled site (Fig. 1). Whether this is the result of technical manipulation of the chromosomes, or a function of chromosomal condensation remains to be determined.

To determine if the $(TTAGGG)_n$ tandem repeats are directly at the ends of human chromosomes, high molecular weight DNA was digested with Bal 31 nuclease for increasing amounts of time (12). This enzyme progressively shortens DNA molecules from their ends , and hence sequences that are at the original chromosome termini will be progressively shortened, while internal DNA sequences will be unaffected by moderate digestion. Genomic DNA sequences complementary to clone pHuR 93 are devoid of most restriction enzyme recognition sites and remain as high molecular weight fragments following Sau 3AI or Rsa I digestion and gel electrophoresis. Digestion with Bal 31 nuclease prior to restriction enzyme digestion, however, shows a progressive shortening and eventual loss of over 99% of the genomic DNA sequences complementary to clone pHuR 93 (12). In contrast, genomic DNA sequences complementary to either an Alu repetitive sequence probe or a chromosome 16 specific centromeric repeat sequence are unaffected by Bal 31 digestion (12). The observed kinetics of Bal 31 digestion (200 bp/min) is consistent with our estimate that 250 to 1000 hexamers are present at each human telomere. The complete disappearance of the telomeric $(TTAGGG)_n$ sequences occurs after the removal of approximately 4000 base pairs.

In order to use in situ hybridization to define the evolutionary origin of the human telomere, synthetic oligodeoxynucleotides were used to determine the sequence dependence for cross-hybridization of the $(TTAGGG)_n$ repeat. We

synthesized oligomers of the known telomeric repeats of human, Tetrahymena, Paramecium, Oxytricha, Saccharomyces and Arabidopsis, as well as the closely related repetitive sequences $(GGTA)_n$ and $(GGGTA)_n$, reported from the crab and Physarum genomes, respectively (Table IA). These oligomers were used in various annealed combinations for thermal denaturation analysis (17).

The human telomere sequence was found to cross-hybridizes with both the Tetrahymena and Paramecium sequences, even though there is a base-mismatch every six nucleotides in these duplexes. In 50 mM NaCl, the melting temperature of the human-Tetrahymena complex is more stable than the human-Paramecium complex, presumably since G-T base mismatches are more stable than T-T base mismatches. The thermal stability of the human-Tetrahymena complex is only 5°C lower than the human $(TAACCC)_7 \cdot (GGGTTA)_7$ complex (Table IB). Interestingly, the human telomere sequence $(TAACCC)_7$ also formed complexes with the plant $(GGGTTTA)_6$, yeast $(TGTGTGGG)_5$, and Oxytricha $(GGGGTTTT)_5$ telomere sequences, as well as the related repetitive sequences $(GGTA)_{10}$ and $(GGGTA)_8$, although the duplexes were 10-28°C less stable and exhibited lower hyperchromicities than the human-Paramecium telomere complexes (Table IB). The ability of these divergent telomere sequences to form stable hydrogen-bonded complexes under physiologically relevant conditions may explain their functional interchangeability as telomeres in yeast artificial chromosomes (26), and as primers for telomerase activity (6).

To define conditions under which the human sequence would not cross-hybridize to other telomeric sequences, thermal denaturations in various concentrations of salt and formamide were conducted with the most stable human-Tetrahymena and human-Paramecium complexes. It is not expected that other sequences with single-base changes from the human $(TTAGGG)_n$ sequence would form more stable mis-matched duplexes than these sequences (24,25). In 0.4 M NaCl,

30% formamide, the cross-species complexes are less stable than in normal salt solutions (17). The human-Tetrahymena complex, in particular, exhibits a dramatic reduction in thermal stability, and a decreased hyperchromicity.

When in situ hybridization was conducted in 0.4 M NaCl, 30% formamide at 37°C, some cross-hybridization of the Tetrahymena $(GGGGTT)_7$ sequence was observed on human telomeres. The fluorescent signal intensity was, however, considerably weaker and less consistent than the signal obtained with the human sequence $(GGGTTA)_7$. The difference between these hybridization signals was quite apparent, and hence metaphase spreads from 91 different species (17), including representative orders of bony fish, reptiles, amphibians, birds, and mammals, were hybridized to biotinylated $(GGGTTA)_7 \cdot (TAACCC)_7$ using this hybridization protocol. A final "melt" wash, in which the temperature was raised from 37°C to 50°C (in 33 mM NaCl) was used for at least one species of each order, to confirm the identity of the $(GGGTTA)_7$ hybridization (17). None of the other synthetic telomere sequences cross-hybridize to vertebrate telomeres under the "melt" wash conditions. In all 91 species, hybridization to the telomeres of all chromosomes was observed, regardless of chromosome number or size (17).

While the number of species studied is a small percentage of the total number of extant vertebrate species, it represents a diverse range of this subphylum. It is unlikely, therefore, that if a variety of telomere sequences were utilized by vertebrates, evidence of this variation would not have been detected in this study. The human sequence does not cross-hybridize to insect or plant telomeres under stringent hybridization conditions, however, indicating that telomere sequence evolution has occurred during animal evolution. The average rates of DNA sequence evolution differ between taxonomic groups, varying between 0.25 and 1.25 percent per million years (27).

The conservation of the $(TTAGGG)_n$ sequence and its telomeric location, in species thought to share a common ancestor over 400 million years ago (17), strongly suggest this molecular "fossil" (4) is the functional vertebrate telomere.

The recent finding that telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs (28) has suggested that telomere function may involve novel DNA-DNA or DNA-protein conformations and interactions (3,28,29). The $(TTAGGG)_n$ repeat, present at the telomeres of vertebrate chromosomes (12,17) can prime, in vitro, the addition of tetrahymena specific $(TTGGGG)_n$ repeats when the telomere terminal transferase of tetrahymena is used (6). We have shown recently that the $(TTAGGG)_n$ sequence can function as a telomere in yeast, allowing the cloning of large (50-250 kb) human telomeric DNA fragments as yeast artificial chromosomes (30). The end point of the human physical and genetic maps has been obtained.

References

1. M.F. Singer, Int. Rev. Cytol 76, 67-112 (1982).
2. A.M. Weiner, P.L. Deininger, and A. Efstratiadis, Ann. Rev. Biochem. 55, 631-661 (1986).
3. E.M. Blackburn and J.W. Szostak, Ann. Rev. Biochem. 53, 163-194 (1984).
4. A.M. Weiner, Cell 52, 155-157 (1988).
5. C.W. Greider and E.H. Blackburn, Cell 43, 405-413 (1985).
6. C. W. Greider and E.H. Blackburn, Cell 51, 887-898 (1987).
7. E. Henderson, C.C. Hardin, S.K. Walk, I. Tinoco, Jr., and E.H. Blackburn, Cell 51, 899-908 (1987).
8. C.W. Greider and E.H. Blackburn, Nature 337, 331-337 (1989).
9. A.M. Zahler and D.M. Prescott, Nucl. Acids Res. 16, 6953-6972 (1988).

10. G.B. Morin and T.R. Cech, Cell 46, 873-883 (1986).

11. A.F. Pluta and V.A. Zakian, Nature 337, 429-433 (1989).

12. R.K. Moyzis, J.M. Buckingham, L.S. Cram, M. Dani, L.L. Deaven, M.D. Jones, J. Meyne, R.L. Ratliff, and J.-R. Wu, Proc. Natl. Acad. Sci. USA 85, 6622-6626 (1988).

13. R.K. Moyzis, K.L. Albright, M.F. Bartholdi, L.S. Cram, L.L. Deaven, C.E. Hildebrand, N.E. Joste, J.L. Longmire, J. Meyne, and T. Schwarzacher-Robinson, Chromosoma 95, 375-386 (1987).

14. A. Rich, A. Nordheim, and A. H.-J. Wang, Ann. Rev. Biochem, 53, 791-846 (1984).

15. E.H. Blackburn and P.B. Challoner, Cell 36, 447-457 (1984).

16. L.H.T. Van dear Ploeg, A.Y.C. Liu, and P. Borst, Cell 36, 459-468 (1984).

17. J. Meyne, R.L. Ratliff, and R.K. Moyzis, Proc. Natl. Acad. Sci., in press (1989).

18. A. Baroin, A. Prat, F. Caron, Nucl. Acids Res. 15, 1717-1728 (1987).

19. L.A. Klobutcher, M.T. Swanton, P. Donini, D.M. Prescott, Proc. Natl. Acad. Sci. USA 78, 3015-3019 (1981).

20. J. Shampay, J.W. Szostak, E.H. Blackburn, Nature 310, 154-157 (1984).

21. E.J. Richards and F.M. Ausubel, Cell 53, 127-136 (1988).

22. D.M. Skinner, W.G. Beattie, F.R. Blattner, B.P. Stark, J.E. Dahlberg, Biochemistry 13, 3930-3937 (1974).

23. E.M. Johnson, Cell 22, 875-886 (1980).

24. D.M. Gray and R.L. Ratliff, Biopolymers 16, 1331-1342 (1977).

25. D.M. Gray, M. Vaughan, and R.L. Ratliff, Nucl. Acids Res. 8, 3695-3707 (1980).

26. D.T. Burke, G.F. Carle, and M.V. Olson, Science 236, 806-812 (1987).

27. R.J. Britten, Science 231, 1393-1398 (1986).

28. E. Henderson, C.C. Hardin, S.K. Walk, I. Tinoco, Jr., and E.H. Blackburn, Cell 51, 899-908 (1987).

29. Y. Oka and C.A. Thomas, Jr., Nucleic Acids Res. 15, 8877-8898 (1987).

30. H.C,. Riethman, R.K. Moyzis, J. Meyne, D.T. Burke, and M.V. Olson, Proc. Natl. Acad. Sci. in press (1988).

Table IA - Synthetic Repetitive DNA Oligodeoxynucleotides

Oligomers	Organism	T_m (°C)	Reference
$(GGGTTA)_7 \cdot (TAACCC)_7$	Human	70	12
$(GGGGTT)_7 \cdot (AACCCC)_7$	Tetrahymena	74	15
$(GGGTTT)_7 \cdot (AAACCC)_7$	Paramecium	72	18
$(GGGGTTTT)_5 \cdot (AAAACCCC)_5$	Oxytricha	68	19
$(TGTGTGGG)_5 \cdot (CCCACACA)_5$	Saccharomyces	70	20
$(GGGTTTA)_6 \cdot (TAAACCC)_6$	Arabidopsis	68	21
$(GGTA)_{10} \cdot (TACC)_{10}$	Crab	68	22
$(GGGTA)_8 \cdot (TACCC)_8$	Physarum	70	23

Oligodeoxynucleotides of 40 to 42 nucleotide lengths were synthesized, hybridized and denatured in 5.0 mM NaCl (12,17). T_m temperature is taken at the last linear height increase in hyperchromicity (17).

Table IB – Melting Temperatures of Mismatched Oligodeoxynucleotide Complexes

Oligomers	Organisms	T_m (°C)
$(GGGGTT)_7 \cdot (TAACCC)_7$	Tetrahymena-Human	65
$(GGGTTT)_7 \cdot (TAACCC)_7$	Paramecium-Human	60
$(GGGTTTA)_6 \cdot (TAACCC)_7$	Arabidopsis-Human	50
$(GGGTA)_8 \cdot (TAACCC)_7$	Physarum-Human	50
$(GGGGTTTT)_5 \cdot (TAACCC)_7$	Oxytricha-Human	38
$(TGTGTGGG)_5 \cdot (TAACCC)_7$	Saccharomyces-Human	34
$(GGTA)_{10} \cdot (TAACCC)_7$	Crab-Human	32

Figure Legends

Fig. 1. In situ hybridization of biotin-labeled $(GGGTTA)_7 \cdot (TAACCC)_7$ oligomers to human metaphase chromosomes (12,17).